

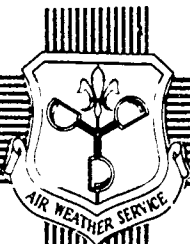


USAFETAC/PR-90/006

AD-A240 489



2



SHORT-TERM HOURLY TEMPERATURE INTERPOLATION

by

Major Walter F. Miller



DTIC
SELECTE
SEP 17 1991
S B D

DECEMBER 1990

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION IS UNLIMITED

91-10649



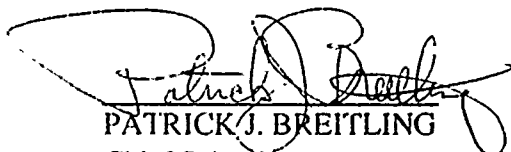
USAF
ENVIRONMENTAL TECHNICAL
APPLICATIONS CENTER

Scott Air Force Base, Illinois, 62225-5438

620 37 6 16

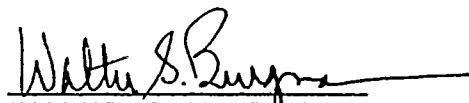
REVIEW AND APPROVAL STATEMENT

USAFETAC/PR-90/006, *Short-Term Hourly Temperature Interpolation*, December 1990, has been reviewed and is approved for public release. There is no objection to unlimited distribution of this document to the public at large, or by the Defense Technical Information Center (DTIC) to the National Technical Information Service (NTIS).



PATRICK J. BREITLING
Chief Scientist

FOR THE COMMANDER



WALTER S. BURGMANN
Scientific and Technical Information
Program Manager
10 December 1990

REPORT DOCUMENTATION PAGE

2. Report Date: December 1990
3. Report Type: Project report
4. Title: Short-Term Hourly Temperature Interpolation
6. Author: Maj Walter F. Miller
7. Performing Organization Name and Address: USAF Environmental Technical Applications Center (USAFETAC), Scott AFB, IL 62225-5438
8. Performing Organization Report Number: USAFETAC/PR-90/006
12. Distribution/Availability Statement: Approved for public release; distribution is unlimited.
13. Abstract: Describes the development of a computer method for interpolating missing hourly temperatures in weather observing records. Although a long-term interpolation method is under consideration, this report deals with replacing missing temperatures over periods of less than 6 consecutive hours.
14. Subject Terms: WEATHER, CLIMATOLOGY, SURFACE TEMPERATURE, INTERPOLATION.
15. Number of Pages: 18
17. Security Classification of Report: Unclassified
18. Security Classification of this Page: Unclassified
19. Security Classification of Abstract: Unclassified
20. Limitation of Abstract: UL

Standard Form 298

PREFACE

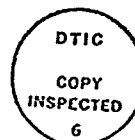
This study documents the results of USAFETAC internal branch tasking (IBT) 6070506, "Fill in of Hourly Temperature for Short-Term Missing Data." The USAFETAC/DNO analyst was Major Walter F. Miller.

The original request (USAFETAC Project 60705, "Request for Engineering Weather Data Support") was to USAFETAC/ECE by HQ USAF/LEEEU for calculation of cooling degree hours (CDH). When ECE discovered that hourly temperature records for many of the locations that required CDH were incomplete, they asked DNO to provide a way to interpolate the missing data. Problems with the records ranged from isolated missing reports at various stations to stations with climatological data (maximum and minimum values) only. Limited-duty stations, of course, do not report at night or on weekends, and synoptic stations report only every 3 hours.

The task of replacing the missing or incomplete data was divided into two parts: (1) a short-term temperature interpolator for less than 6 hours of missing data and (2) a long-term climatic technique to fill in missing temperatures for periods greater than 6 hours. This report--dealing only with the short-term tasking--offers a method for calculating (interpolating) hourly temperatures when less than 6 consecutive hours of data is missing.

CONTENTS

	<i>Page</i>
1. INTRODUCTION	1
1.1 The Tasking	1
1.2 The Programs	1
1.3 The Algorithms	1
2. DATA	3
2.1 Temperature Data	3
2.2 Weather Stations	3
2.3 Temperature Units	3
3. METHODOLOGY	4
3.1 Linear Interpolation	4
3.2 Cubic Polynomial Vs. Cubic Spine Interpolation	4
3.3 Independent Variables	4
3.4 Data Point (Hours) Selection	4
3.5 Predicted Values	4
3.6 Evaluation	5
4. ANALYSIS, RESULTS, AND VERIFICATION	6
4.1 Analysis Summary	6
4.2 Effects of Time Series Length and Gap Length on Performance	6
4.3 Test of Interpolator Performance on 3-Hourly Synoptic Data	7
4.4 Accuracy	8
5. CONCLUSION	9
5.1 RMSE Error	9
5.2 Limitations	9
5.3 Applicability	9
BIBLIOGRAPHY	10
SPECIALIZED TERMS AND ACRINABS	11



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

SHORT-TERM HOURLY TEMPERATURE INTERPOLATION

1. INTRODUCTION

1.1 The Tasking. The tasking requested a method for interpolating hourly temperatures for the calculation of Cooling Degree Hours (CDH). Temperature data from relational database (DB2) tables was used to verify the results. Section 2 of this report identifies the locations chosen. Section 3 describes the methodology of the linear and third-order polynomial interpolators, and Section 4 describes verification of the algorithm. For this verification, data gaps of various sizes were filled, along with nonstandard hours for synoptic stations, which report only every 3 hours. For gaps up to 6 hours, the Root Mean Square Error (RMSE) remains less than 2° F, with a bias less than 0.1° F.

1.2 The Programs. Two Statistical Analysis Systems (SAS) programs were delivered to satisfy the IBT 6070506 requirement. The first (INTERP) uses a cubic polynomial to interpolate missing data points. The second (TFILLER) is used to test the accuracy of the interpolator by providing bias and RMSE statistics between observed and interpolated data. TFILLER also converts hourly observations into 3-hourly observations when requested to do so.

1.3 The Algorithm. The algorithm chosen for this project is designed to interpolate missing temperature values for periods up to 6 consecutive hours. One of the most common interpolation algorithms is linear; however, when working with nonlinear data such as the peak of the temperature curve, a large error can be introduced by using points near the base of the rapid rise in temperature, as shown in Figure 1. A quadratic might be better near the maximum or minimum daily temperature, but it doesn't handle the transition from maximum to minimum temperature. The cubic spline or third-order polynomial interpolation has been supported by various authors (Pielke, 1984, and Koehler, 1977) and handles continuous values such as temperature and pressure accurately.

TEMPERATURE FOR SCOTT AFB, IL

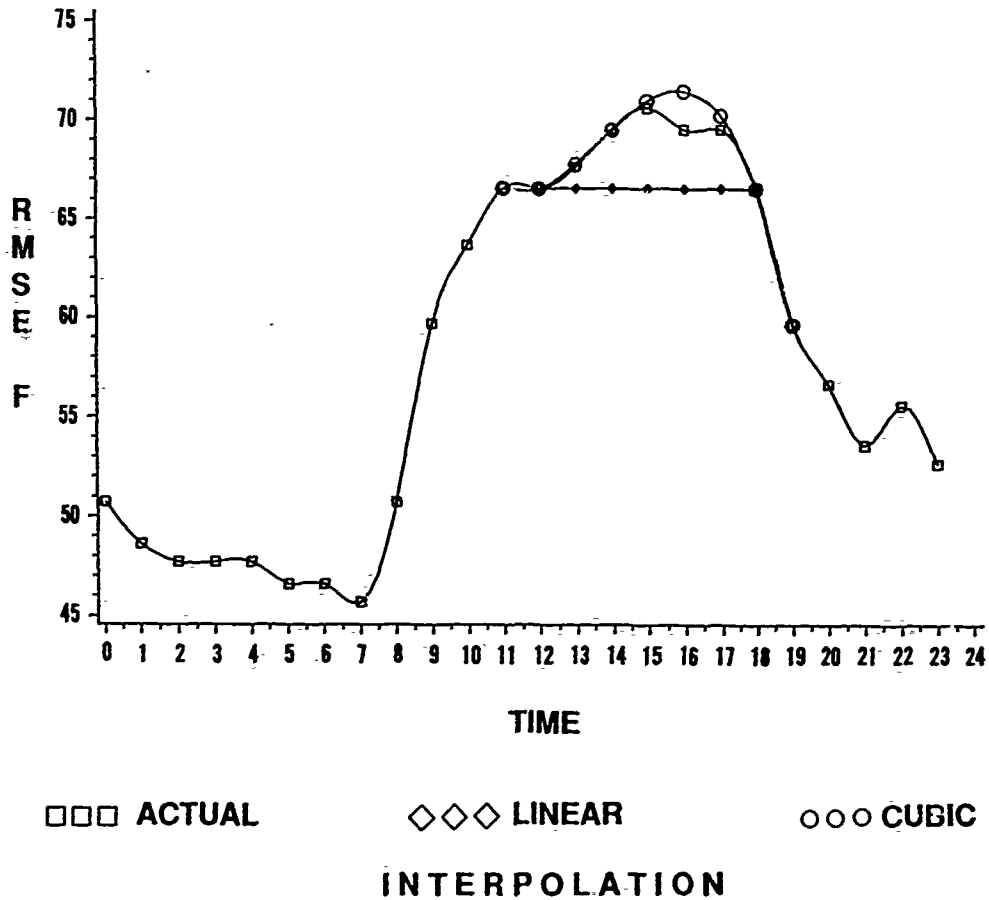


Figure 1. Comparison of Linear and Cubic Spline Interpolation. Five hours of missing data were created between 1300 and 1700Z. For the linear interpolation, one point on either side of the missing data was used. The cubic spline used two points on either side of the data gap.

2. DATA

2.1 Temperature data for the study was from DB2 tables. The data is stored in tenths of degree Fahrenheit for various military and civilian reporting stations. Temperature data is normally available from 1959 to 1988 (except Scott AFB, for which only 1962 to 1985 data is available).

2.2 Weather Stations. USAFETAC/ECE chose the following stations for verification; they are representative of continental, coastal, desert, subtropical, and tropical regions. Clark AB, RP, was on the original list, but the required data was not in the standard DB2 tables.

<u>STATION</u>	<u>WMO #</u>	<u>LAT</u>	<u>LON</u>	<u>ELEV (Meters)</u>
Scott AFB, IL	724338	38° 33' N	89° 51' W	138
Pease AFB, NH	726055	43° 05' N	70° 49' W	31
McChord AFB, WA	742060	47° 09' N	22° 29' W	98
Luke AFB, AZ	722785	33° 32' N	112° 23' W	332
Homestead AFB, FL	722026	25° 29' N	90° 23' W	2
Mildenhall AB, UK	035773	52° 22' N	0° 29' W	10
Ramstein AB, GR	106140	49° 26' N	7° 36' E	238
Honolulu, HI	911820	21° 21' N	157° 56' W	5

2.3 Temperature Units. Temperature was converted from tenths of degrees to whole degrees Fahrenheit before the interpolation began. Specific units are not needed for this algorithm.

3. METHODOLOGY

3.1 Linear interpolation can be used to determine the value of missing temperature for short intervals. Linear interpolation, in its simplest form, is the mean of the two data points on either side of the evenly spaced data gap that one is trying to fill. Linear interpolation produces good results over short time spans and for straight time series. Most meteorological time series are not linear; the diurnal temperature curve, for example, is close to sinusoidal.

3.2 Cubic Polynomial Vs. Cubic Spline Interpolation. Current literature makes a strong case for using a cubic spline interpolation for meteorological variables, and the Statistical Analysis Systems (SAS) currently available at USAFETAC makes this option easy. However, a cubic spline requires two data points evenly spaced on both sides of the gap through which one needs to interpolate. In climatological data, gaps can be of varying lengths, and the data necessary for a cubic spline could be missing. The cubic polynomial has many of the advantages of the cubic spline, but the minimum four data points do not need to be evenly spaced. If additional data is available, that data can be used in the interpolation. For these reasons, a cubic polynomial was selected for this project.

3.3 Independent Variables. A third-order polynomial or cubic can be obtained by using the square and cube of the variable as input. In this way, SAS linear regression can be used to obtain the coefficients of the cubic. Our independent variables for this study are the hour, hour-squared, and hour-cubed. Minutes are converted to a decimal by dividing by 60 and adding the decimal to the current hour; for example, 1030 becomes 10.50.

3.4 Data Point (Hours) Selection. A regression is done on each day. The number of hours used in a regression is controlled by NUMHRS, which can vary from 5 to 15. Five hours (data points) are the minimum needed to do a cubic spline assuming the middle point is missing. As the number of points in a time series increases, the error also increases. Fifteen is chosen as a compromise between the need to fill large data gaps and the increase in error. Another check is performed on the data to be sure there are two valid, "non-missing" observations either side of the center point. This restriction prevents extrapolation from occurring. Each hour of the day becomes the first value of the time series. The start hour is then shifted to zero so that there is no problem crossing into another day and having a discontinuous function in time (e.g., 22, 23, 0, 1).

3.5 Predicted Values. SAS produces a "predicted" value for each hour of the time series. Initially only the center point is used to fill in the value. To increase efficiency, the user controls the number of points that can be filled on each pass, HRINC. The number of points filled cannot be more than the number of hours used. The program takes 10 CPU seconds per hour per year. To fill each hour individually for a 30-year record takes 69 CPU minutes. If you fill 3 hours each pass, the run time drops to 23 CPU minutes.

3.6 Evaluation. The accuracy of the interpolation is evaluated using the bias and RMSE. The bias is defined as the mean of the error of estimate, as shown below:

$$BIAS = \frac{\Sigma(T_o - T_p)}{N}$$

The error of estimate is the difference between the actual observed temperature (T_o) and the predicted temperature (T_p) from the cubic polynomial. N is the number of observations used. A bias near zero indicates that there is an equal number of predicted values above and below the actual value. RMSE is calculated as shown here:

$$RMSE = \sqrt{\frac{\Sigma(T_o - T_p)^2}{N-1}}$$

The RMSE gives a measure of the size of the differences between observed and predicted values; the square of the RMSE is really a measure of the variance between observed and predicted (interpolated) temperatures.

4. ANALYSIS, RESULTS, AND VERIFICATION

4.1 Analysis Summary. For reasons stated in Section 3, cubic polynomials were selected as the functional form for short-term temperature interpolation. Cubics can be fitted to time series of any length greater than or equal to a minimum length. For purposes of interpolating across a data gap, the minimum length of the time series for a gap of length 1 is five points; i.e., one missing point plus two points on either side of the missing point. Accordingly, a minimum time series length of five points is used throughout this project. Unknowns to be determined were the optimum length of the time series to be fitted and the effect of increasing gap size on performance of the interpolator. These tests can be done using an option in the INTERP algorithm that replaces actual data with a missing indicator. The amount of missing data is set by HRINC and centered in the time series.

4.2 Effects of Time Series Length and Gap Length on Performance. In determining the most satisfactory short-term temperature interpolator, three principal efforts were made. First, tests were conducted to determine the maximum length of time series to use in determining the cubic polynomial (see 4.2.1). Second, with the maximum time series length determined, tests were conducted over a range of lengths (up to the maximum determined in the first test) regarding the effect on interpolation accuracy of varying the data gap from a minimum of 1 hour missing (five points, including the gap) to a maximum of 11 hours missing (15 points, including the gap) (see 4.2.2). On the basis of the varying gap length tests, the length of the gap was limited to 6 missing hours. Third, a special study was done of the 6-missing-hour case, in which performance of the 6-hour gap interpolator was studied as a function of the length of the cubic, from a minimum of 10 points to a maximum of 15, including missing points (see 4.2.3).

4.2.1 The first set of tests are run to determine how long a time series can be used in a cubic polynomial and still predict temperature accurately. In these tests, the number of hours in the time series varied from 5 to 19. The center value was removed and replaced with an interpolated value. Temperature data for Scott AFB for year 1975 were used for this test. Figure 2 shows RMSE as a function of the number of hours as the lowest line. When the number of hours equals 5, a true cubic spline is performed. The RMSE increases with the number of hours used and exceeds 2 for more than 17 hours. The bias is less than a thousandth of a degree Fahrenheit for time series under 15 hours. Based on the above results, we limited the length of data submitted to the cubic to 15 hours.

ROOT MEAN SQUARE ERROR FOR SCOTT AFB, IL

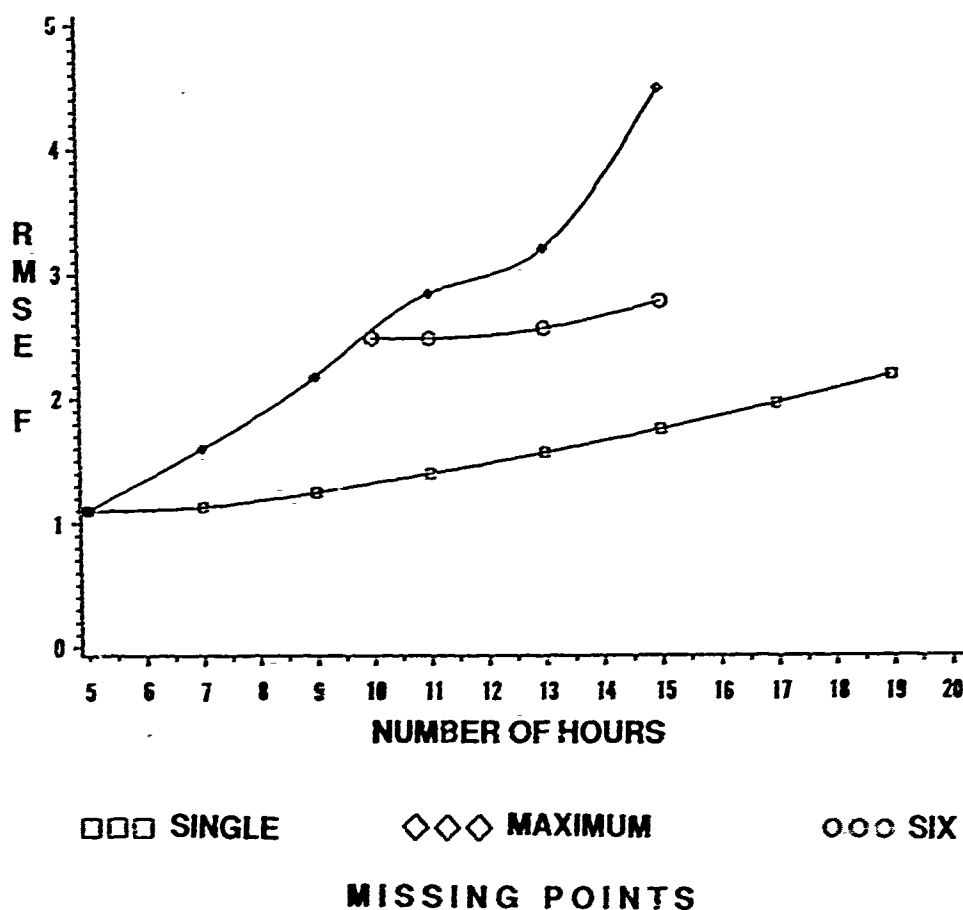


Figure 2. Root Mean Square Error as a Function of Time-Series Length and Number of Missing Hourly Temperatures at Scott AFB.

4.2.2 In the second set of tests, the uppermost line (marked by diamonds in Figure 2) is the RMSE using the same data, but removing all but four points--two at each end of the time series. For example, with 5 data points, 1 hour is set to missing; a maximum of 11 hours are set to missing for 15 data points. The RMSE quickly rises from 1.1 (5 data points) to 4.5°F (15 data points). Similarly, the bias (not shown in Figure 2) increases with the size of the data gap. For more than 11 missing points, the bias is more than a degree.

4.2.3 Using results from the above test, the number of missing observations was limited to six. In Figure 2 these are shown for NUMPRS greater than 11 by the short line marked with circles. With 6 missing hours in a 15-hour time series, the RMSE is 2.8°F. The bias remains a few tenths of degrees, with maximum interpolation error of 24°F.

4.3 Test of Interpolator Performance on 3-Hourly Synoptic Data. To test the interpolation technique, TFILLER was coded in SAS. The purpose of TFILLER is to convert hourly data to 3-hourly data and verify it. When the user requests the 3-hourly option (THREEHR = 1), the temperature for any hour not divisible by three is set to missing. This new temperature array is sent to the interpolation program, and a predicted array is returned. The predicted observations from INTERP is then compared to the actual values. The bias and RMSE were calculated using the difference between actual and predicted temperature.

4.3.1 As mentioned earlier, completing missing hourly observations from a synoptic station would be the more common use for this program. To test its accuracy, 3-hourly data was created from hourly data for the eight stations listed in Section 2. The stations were chosen to be representative of different climatic regimes with significant military operations. The data was processed in a time series of 13 hours, with 3 hours being predicted during each regression even if a fill was not required. We chose 13 hours as the length of the time series because it would include 5 actual observations (00, 03, 06, 09, and 12Z). If one of these observations were missing, we still met the minimum requirements to do the cubic polynomial: a gap less than 6 hours and a minimum of two data points at or either side of the center. Ten is the minimum number of hours required in the time series to have four 3-hourly observations.

4.3.2 The table below shows the results of this test. RMSE is between 1 and 2 for the entire data set. In reviewing monthly RMSE, a few isolated values less than 1 or greater than 2 occur. Luke AFB, in the desert, has the poorest overall RMSE. The bias for all the stations still remains small, at less than a tenth of a degree Fahrenheit. One percent of the errors between observed and interpolated values were greater than the 99th percentile and 1 percent were less than the 1st percentile, as shown. The maximum error between the observed and interpolated for any hour is 17.3°F at Luke, but the maximum error for other stations is in the low teens. With 13 hours in the time series, 2 hours will be filled during each pass for a synoptic station using five equally spaced data points for the regression. A gap of 5 hours may occasionally occur in the data when one synoptic observation is missing. If more than one synoptic observation is missing, however, an interpolation cannot be performed.

Comparison of Interpolated Temperature using 3-Hourly Data to Actual Observations with a 2-Hour Gap.

Station Name	Block Number	Bias	RMSE	Percentile	
				99	1
Scott AFB, IL	724338	0.002	1.44	3.64	-4.07
Luke AFB, AZ	722785	-0.006	1.99	4.67	-5.58
McChord AFB, WA	742060	-0.017	1.31	3.33	-3.70
Pease AFB, NH	726055	-0.002	1.27	3.33	-3.41
Mildenhall AB, UK	035773	-0.004	1.20	3.01	-3.22
Ramstein AB, GR	106140	-0.010	1.44	3.70	-4.05
Homestead AFB, FL	722026	0.038	1.36	3.59	-3.85
Honolulu, HI	911820	-0.072	1.20	3.07	-3.40

4.4 Accuracy. The AWS Capabilities Master Plan (AWS/CMP-89/001) lists the accuracy of the current Air Force temperature sensor, the TQM-11, as 1.0°F. Its replacement, the FMQ-8, is also accurate to 1.0°F. The RMSE results from the tests, then, are only slightly larger than the accuracy of the equipment.

5. CONCLUSION

5.1 RMSE Error. USAFETAC/DNO has developed an effective temperature interpolation scheme using a cubic polynomial. When interpolating gaps of less than 6 hours, the RMSE is below 2.8°F. These errors are randomly distributed between being over- and under-estimates of the temperature. For synoptic data, the RMSE error remains less than 2°F and has a low bias.

5.2 Limitations. Constraints are placed in the software to limit a data gap to 6 hours and a time series to 15 hours. These restrictions are chosen to prevent large interpolation errors. There must be at least two observations on each side of the middle point to keep extrapolation errors from occurring. An option in the SAS program, TFILLER, allows the user to test the accuracy of the interpolation scheme for stations with hourly observations.

5.3 Applicability. The interpolator should be applicable to any other continuous atmospheric variable such as pressure, dew-point temperature, or density altitude. The user should run a test similar to the one described in this report before proceeding; the software is designed to do this for any variable in the DB2 tables.

BIBLIOGRAPHY

AWS Capabilities Master Plan, AWS/CMP-89/001, HQ Air Weather Service, Scott AFB, IL, August 1989.

Koehler, T.L., *A Test of Seven Methods which Perform Grid to Observation Interpolations, in Meteorological Applications of Satellite Indirect Soundings II*, Project Report, Department of Meteorology, University of Wisconsin-Madison, Madison, WI, December 1977.

Pielke, R.A., *Mesoscale Meteorological Modeling*, Academic Press, Orlando, FL, 1984.

SPECIALIZED TERMS AND ACRINABs

ACRINAB	acronym, initialism, abbreviation
CDH	cooling degree hour
CPU	computer processing unit
DB2	IBM relational database in use at USAFETAC
FMQ-8	AWS temperature/dew point set
HRINC	The number of hours filled during each step.
INTERP	SAS software that performs a cubic polynomial interpolation.
NUMHRS	The number of hours in a time series.
RMSE	Root Mean Square Error
SAS	Statistical Analysis Systems
TFILLER	SAS software that converts an hourly reporting station to a 3-hourly reporting station and performs verification.
TMQ-11	AWS humidity/temperature measuring set
Z	Zulu time

DISTRIBUTION

AWS/DO, Scott AFB, IL 62225-5008	2
AWS/XT, Scott AFB, IL 62225-5008.....	2
OL A, HQ AWS, Buckley ANG Base, Aurora, CO 80011-9599.....	1
SD/CWDA, PO Box 92960, Los Angeles, CA 90009-2960	1
OL-K, HQ AWS, NEXRAD Opnl Support Facility, 1200 Westheimer Dr., Norman, OK 73069	1
OL-L, HQ AWS, Keesler AFB, MS 39534-5000.....	1
OL-M, HQ AWS, McClellan AFB, CA 95652-5609	1
Det 1, HQ AWS, Pentagon, Washington, DC 20330-6560.....	3
Det 2, HQ AWS, Pentagon, Washington, DC 20330-5054.....	3
Det 3, HQ AWS, PO Box 3430, Onizuka AFB, CA 94088-3430.....	1
Det 9, HQ AWS, PO Box 12297, Las Vegas, NV 89112-0297	1
1WW/DN, Hickam AFB, HI 96853-5000.....	3
11WS/DON, Elmendorf AFB, AK 99506-5000	1
20WS/DON, APO San Francisco 96328-5000	1
30WS/DON, APO San Francisco 96301-0420	1
2WW/DN, APO New York 09094-5000.....	3
7WS/DON, APO New York 09403-5000	1
28WS/DON, APO New York 09127-5000.....	1
31WS/DON, APO New York 09136-5000.....	1
3WW/DN, Offutt AFB, NE 68113-5000.....	6
9WS/DON, March AFB, CA 92518-5000.....	1
24WS/DON, Randolph AFB, TX 78150-5000.....	1
26WS/DON, Barksdale AFB, LA 71110-5002.....	1
4WW/DN, Peterson AFB, CO 80914-5000	3
2WS/DON, Andrews AFB, MD 20334-5000	20
5WW/DN, Langley AFB, VA 23665-5000.....	3
1WS/DON, MacDill AFB, FL 33608-5000.....	1
3WS/DON, Shaw AFB, SC 29152-5000	1
5WS/DON, Ft McPherson, GA 30330-5000	1
25WS/DON, Bergstrom AFB, TX 78743-5000.....	1
AFGWC/SDSL, Offutt AFB, NE 68113-5000.....	3
USAFETAC, Scott AFB, IL 62225-5438.....	6
7WW/DN, Scott AFB, IL 62225-5008.....	3
6WS/DON, Hurlburt Field, FL 32544-5000	1
15WS/DON, McGuire AFB, NJ 08641-5002	1
17WS/DON, Travis AFB, CA 94535-5986	1
3350 TCHTG/TTGU-W, Stop 62, Chanute AFB, IL 61868-5000.....	2
3395 TCHTG/TTKO, Keesler AFB, MS 39534-5000	2
AFIT/CIR, Wright-Patterson AFB, OH 45433-6583	1
AFCSA/SAGW, Washington, DC 20330-5000	1
NAVOCEANCOMDET, Federal Building, Asheville, NC 28801-2723.....	1
NAVOCEANCOMDET, Patuxent River NAS, MD 20670-5103	1
NAVOCEANCOMFAC, Stennis Space Ctr, MS 39529-5002 (NO DSs, clim sums, etc!!!)	1
NAVOCEANCOMFAC, NAS North Island, San Diego, CA 92135-5130	1
COMNAVOCEANCOM, Code N312, Stennis Space Ctr, MS 39529-5000.....	1
COMNAVOCEANCOM (Capt Brown, Code N332), Stennis Space Ctr, MS 39529-5001.....	1
NAVOCEANO, Code 9220 (Tony Ortolano), Stennis Space Ctr, MS 39529-5001(NO DSs, ETC!!)	1
Maury Oceanographic Library (NOC), Code XJL, Stennis Space Ctr, MS 39529-5001.....	1
FLENUMOCEANCEN, Monterey, CA 93943-5006.....	1
NOARL West, Monterey, CA 93943-5006.....	1